Human β-casein

BY RAE GREENBERG AND MERTON L. GROVES

Eastern Regional Research Center, Philadelphia, Pennsylvania 19118, USA*

SUMMARY. Human β-casein occurs in multiphosphorylated forms having the same amino acid composition but with 0–5 phosphate groups/molecule. Sequence analysis was used to determine whether each of the phosphorylated forms is a mixture of species having a certain number of phosphate groups randomly distributed or whether each form contains phosphate groups on specific seryl or threonyl residues. It was found that forms containing 2, 4 and 5 phosphate groups/molecule are homogeneous with respect to their phosphorylation sites. The monophosphorylated form, however, is a mixture of equal amounts of species phosphorylated at residues 9 or 10.

The major component of the casein system in mature human milk is a protein similar in size, composition, and electrophoretic mobility to cow (*Bos Taurus*) β-casein (Groves & Gordon, 1970; Nagasawa, Kiyosawa & Kuwahara, 1970). Unlike the cow protein which contains a specific number of phosphate groups depending on the genetic type, human β-casein occurs in multiphosphorylated forms with 0–5 phosphate groups/molecule. Plate 1 compares the polyacrylamide disc-gel electrophoretic patterns of multibanded human β-casein with that of the cow caseins (Groves & Gordon, 1970). The bands represented as TS-, R-, and γ-caseins are now known to be fragments of β-casein probably arising from proteolytic cleavage of the parent molecule. Band number I of human β-casein contains 5 phosphate groups/molecule and corresponds in electrophoretic mobility to the cow counterpart. The phosphate values decrease in integer numbers from bands I to V with band VI representing nonphosphorylated material. Separation and purification of these components were carried out on DEAE-cellulose columns as described earlier (Groves & Gordon, 1970). The isolated proteins exhibited identical amino acid compositions and varied only in phosphate content. Table 1 compares these analyses for human and cow β-caseins; each contains about 209 residues/molecule and the 2 proteins differ in only 35 residues.

The phosphate groups

In the light of the above information it was interesting to determine whether each of the multiphosphorylated forms represented a mixture of species having a certain number of phosphate groups randomly distributed or whether each form contained phosphate groups on specific seryl or threonyl residues. Sequence analysis carried out on the intact molecule using a liquid-phase spinning cup sequencer provided the answer (Greenberg, Groves & Peterson, 1976). The sequence of the first 28 residues of the phosphate-free forms of human β-casein is shown in Fig. 1. At residues 5 through 12 is the sequence cluster corresponding to the site of phosphorylation in

* Agricultural Research, Science and Education Administration, US Department of Agriculture.

Table 1. *Amino acid composition of β-caseins (residues/molecule)*

	Human	Cow
Asx	11	9
Thr	9	9
Ser	9	16
Glx	39	39
Pro	39	35
Gly	3	5
Ala	7	5
Val	19	19
Met	3	6
Ile	13	10
Leu	26	22
Tyr	7	4
Phe	5	9
Trp	1	1
Lys	11	11
His	5	5
Arg	3	4

1 5 10
NH_2-Arg-Glu-Thr-Ile-Glu-Ser-Leu-Ser-Ser-Ser-

11 15 20
Glu-Glu-[illegible]-Ile-Pro-Glu-Tyr-Lys-Gln-Lys-

21 25
Val-Glu-Lys-Val-Lys-His-Glu-Ser-

Fig. 1. Amino acid sequence of the first 28 residues of the phosphate-free form of human β-ceasin.

cow β- (residues 13–21) and α_{s1}-caseins (residues 63–70) (Whitney *et al.* 1976) and the probable site in cow α_{s2}- and buffalo β-caseins (Brignon *et al.* 1977; Addeo, Mercier & Ribadeau Dumas, 1977). To identify the phosphorylated sites and quantitate the phosphate, either the entire sequencer output at each seryl or threonyl residue or the aqueous layer from the PTH conversion step was subjected to microphosphorus analysis (Greenberg *et al.* 1976). Although the sequencer yield decreases after passing through an area containing phosphoserine, continued identification of sequential residues is still possible. Basing the calculation on yields obtained before and after such an area, phosphorus recoveries averaged 90% of theoretical.

According to Mercier, Grosclaude & Ribadeau Dumas (1971), the configuration Thr/Ser-X-Glu/SerP (where X is any amino acid) is required for casein phosphorylation. Those residues at position n which are followed by glutamic acid at $n+2$ are termed primary sites and those with phosphoserine at $n+2$, secondary sites. From Fig. 1, the amino terminal portion of human β-casein contains primary phosphorylation sites at positions 3, 9, and 10 and secondary sites at positions 6 and 8. Fig. 2 shows the results of sequence analysis and phosphate determination for the 1, 2, 4, and 5-P forms of the protein. The finding of phosphorus only on specific seryl and threonyl residues in forms phosphorylated at different levels indicates that the components are indeed homogeneous with respect to phosphorylated sites and not merely charge class mixtures containing 2–5 phosphate groups randomly distributed. The location of phosphorus on seryl residues 9 and 10 in the biphosphorylated form suggested that position 10 would be phosphorylated first, this being mediated by the highly acidic nature of positions 11 and 12 (-Glu-Glu). Stepwise phosphorylation

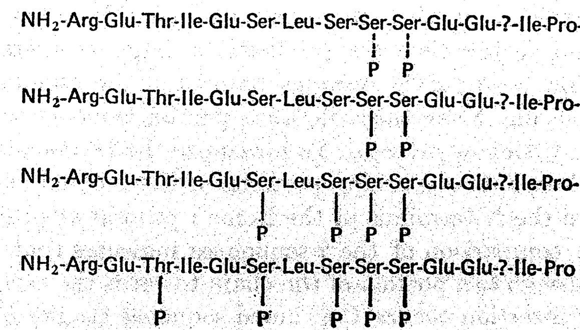

Fig. 2 Location of phosphate groups on the 1, 2, 4, and 5-P forms of human β-casein.

back to serine 6 then might occur. The threonyl residue at position 3, although a primary site, seems to be phosphorylated at a slower rate than the primary or secondary seryl sites. A recent examination of the monophosphorylated human β-casein (Plate 1, band V) revealed that it alone is a mixture, with half the molecules phosphorylated at position 9 and half at position 10. This was determined by quantitating the phosphorus and finding equal amounts on each of the 2 residues. Therefore, in this case only these 2 sites are equivalent; each is followed at $n+2$ by glutamic acid, and the enzyme cannot distinguish between serines 9 and 10. After this, for the 2, 3, 4, and 5-P forms, phosphate addition appears to be stepwise and there are no intermediate forms. Possible phosphate migration after incorporation seems to be ruled out by its absence on the neighbouring seryl residue 8 in the biphosphorylated molecule.

Sequence strategy

Further characterization of the sequence of human β-casein involved cleavage with cyanogen bromide (Groves & Greenberg, 1978). The expected 4 peptides arising from cleavage of the 3 methionyl bonds were fractionated by gel filtration on Bio-Gel P30 followed by chromatography on DE-32 DEAE-cellulose. The purity of these peptides, CB-1 through 4, as well as those described below was monitored by polyacrylamide-disc gel electrophoresis at pH 4·3 and 9·5 in 7·5% (w/v) gels with and without urea. High-voltage paper electrophoresis at pH 6·4 was also used both as an analytical and preparative tool for several small fragments. The sum of the 4 CB-peptides, which contain 78, 16, 37, and 77 residues, respectively, agrees quite well with the composition of the intact molecule (208 *vs.* 210).

Peptides were also produced by tryptic digestion of the protein after masking of the lysyl residues with *S*-ethyl trifluoroacetate according to Goldberger (1967). The enzyme digest after cold piperidine deblocking was subjected to gel filtration on Sephadex G-25 followed by chromatography on phosphocellulose at pH 4·0 using stepwise elution with Na acetate from 0·01 to 0·1 M. In this case, 4 fragments were isolated whose composition accounted for the total amino acid content of human β-casein. Since the protein contains 3 arginine residues, one of which is N-terminal, only 3 peptides would be expected from tryptic digestion of the trifluoroacetylated derivative. It is possible that the extra fragment is the result of a non-specific cleavage by trypsin or the failure to protect a lysine residue. These 4 fragments are 45, 50, 76, and 40 residues in length.

The information provided by the aforementioned peptides plus other purified fragments such as nonspecific cleavage products, a large temperature sensitive fraction, and the amino terminal 28 residues permitted the alignment of all the fractions along the molecule. From this a detailed partial sequence was constructed and comparisons made with cow β-casein. To maximize the homologies between the amino terminal and phosphate-containing portions of the 2 equal-length β-caseins, it is necessary to place the N-terminus of the human protein at position 10 of the cow sequence. Further comparison of the 2 sequences indicates that this 10 residue 'frame shift' carries through to a portion of the chain towards the carboxyl terminus where a compensating insertion occurs. Continued sequence studies now in progress should confirm this observation in the near future.

DISCUSSION

The reason for the phenomenon of multiphosphorylation, earlier thought to be limited to human casein and perhaps mouse or rat milk proteins (McKenzie & Larson, 1978), is unclear. Possible explanations involved single or multiple kinase or phosphatase activity during biosynthesis or a variability in transport mechanism between cow β-casein and the 6 forms of the human counterpart. Recent reports indicate that cow α_{s2}-, α_{s3}-, α_{s4}, and α_{s6}-caseins are a group of proteins (now designated the α_{s2}-complex) having identical peptide chains with differing phosphate content (Brignon *et al.* 1977). This may be analogous to the aforementioned multiphosphorylation although it is not yet clear whether each α_{s2}-casein component is phosphorylated at specific sites. As more of these systems in milks of other species are characterized there will be more opportunities of answering the questions posed by these biosynthetic processes.

The appearance of quantitative variation among the 6 β-casein components is another area worthy of mention. The pattern presented in Plate 1 where bands II and IV (4-P and 2-P forms, respectively) show greater intensities than the others is quite common locally for many individual and most pooled milk samples. Differing band intensities have been ascribed by several groups to genetic control (Voglino, Caone & Ponzone, 1975; Tsuda, 1976). During our screening studies of individual samples many variations in band intensity patterns have been encountered. For example, the β-casein of one donor showed a marked increase in intensity of low phosphate forms over the Plate 1 type pattern. Although this milk was absolutely normal in appearance it was characterized by high pH values routinely ranging from 7·4 to 7·8. There is no evidence linking the preceding 2 observations but other factors such as stage of lactation, health of the mother, or altered chemistry of non-protein milk components such as Ca and phosphorus may play a role in addition to the genetic factor in governing quantitative variations among the 6 human β-casein components.

REFERENCES

ADDEO, F., MERCIER, J.-C. & RIBADEAU DUMAS, B. (1977). *Journal of Dairy Research* **44**, 455.

BRIGNON, G., RIBADEAU DUMAS, B., MERCIER, J.-C., PELISSIER, J. P. & DAS, B. C. (1977). *FEBS Letters* **76**, 274.

GOLDBERGER, R. F. (1967). *Methods in Enzymology* **11**, 317.

GREENBERG, R., GROVES, M. L. & PETERSON, R. F. (1976). *Journal of Dairy Science* **59**, 1016.

GROVES, M. L. & GORDON, W. G. (1970). *Archives of Biochemistry and Biophysics* **140**, 47.

GROVES, M. L. & GREENBERG, R. (1978). *Biochemical Journal* **169**, 337.